Tachyum Prodigy

The World's First Universal Processor

- Tachyum is developing the industry's first Universal Processor, AI, and supercomputing chip – Prodigy
- Prodigy has up to 21x higher AI performance and up to 10x better AI performance per watt than its competition
- Prodigy solves key issues plaguing today's data centers, including high power consumption, low server utilization, and the processor performance plateau that is limiting performance

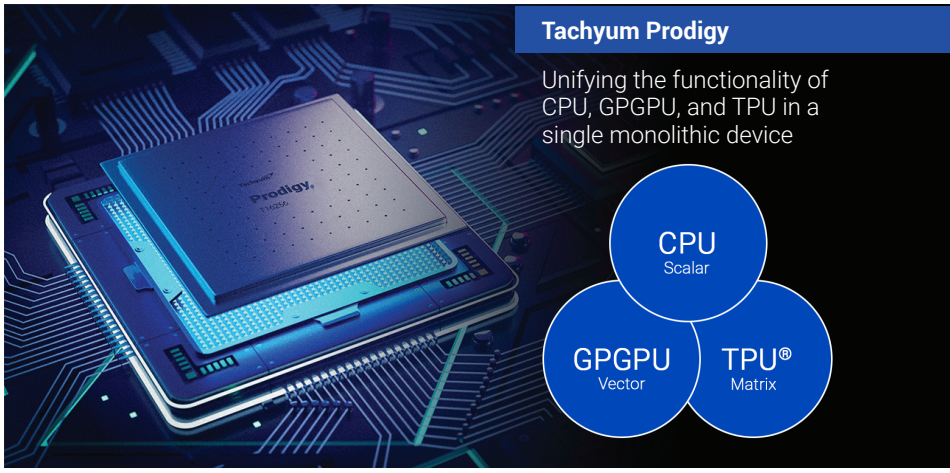
Company and Product Overview

Tachyum is a semiconductor company developing the world's first Universal Processor, Prodigy, which unifies the functionality of CPU, GPGPU, and TPU into a single device, delivering unprecedented performance, power efficiency, and TCO reduction for a wide range of applications and workloads, including cloud, AI, and HPC.

Prodigy eliminates the need for costly and power-hungry accelerators, enabling high-performance data centers to be deployed with a homogeneous architecture, enabling a simple software model, easy maintenance, and the ability to run servers 24/7 by running cloud workloads during peak hours and AI workloads during off hours.

In addition to running its native instruction set architecture, Prodigy runs the binaries for x86, Arm, and RISC-V, providing fast, easy, out-of-the-box testing and evaluation.

Tachyum has approximately 100 employees with engineering teams in Silicon Valley, Slovakia, Taiwan, and the corporate office in Las Vegas, Nevada.



- HPC, Big AI
- Exascale Supercomputers
- Physical AI
- Cloud/ Hyperscale
- Big Data, Analytics
- Databases, Storage
- Edge/ Telco
- Crypto/Digital Currency

Target Markets and SKUs

The Prodigy family of processors encompasses multiple product SKUs ranging from 1024 to 32 cores with a wide range of performance, power, and features to address a wide array of important markets. The markets and SKUs are presented inside, highlighting the Prodigy Series flexibility and ability to excel in a broad array of applications and workloads.

Key Features

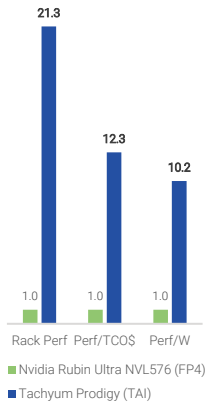
Features	Benefits
1024 64-bit cores running at 6 GHz	Highest performance for compute workloads
HW coherency supports up to 16-socket systems	High scalability for powerful compute nodes
24 DDR5-17600 memory controllers	High memory capacity and bandwidth for LLMs
6.75 TB/s - 13.5 TB/s* of memory bandwidth	*Bandwidth amplification doubles bandwidth
128 lanes of PCIe 7.0 with 64 controllers	High performance NICs, large NVMe storage arrays
Runs native and x86, Arm, and RISC-V binaries	Fast, easy, out-of-the-box testing and evaluation
Advanced matrix and vector units	High-performance AI and HPC
FP64, FP32, TF32, BF16, Int8, FP8, FP4, TAI data types	Converged, homogeneous data centers
Sparsity	Maximum AI performance and memory efficiency

Performance

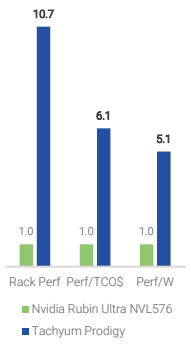
Prodigy vs. NVIDIA Rubin Ultra AI Rack Performance

Inference

Prodigy TAI vs. Rubin Ultra FP4 (Inference)

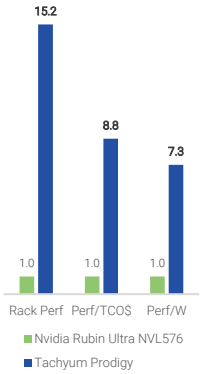


Prodigy vs. Rubin Ultra FP4 (Inference)



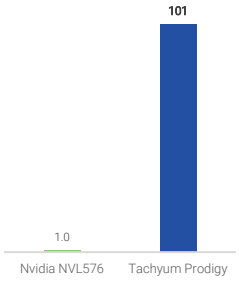
Training

Prodigy vs. Rubin Ultra FP4 (Training)



Cabinet Memory Capacity

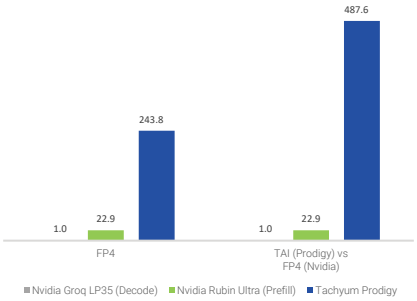
Maximum Rack Memory Prodigy vs. Rubin Ultra



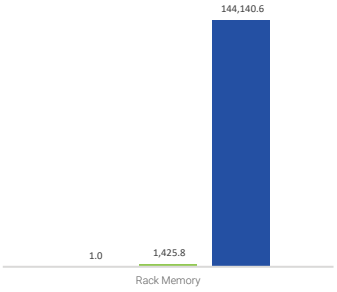
Prodigy provides up to **21.3x higher AI Rack Performance** and **>100x higher Cabinet Memory Capacity** than Rubin Ultra NVL576

Prodigy High-End Rack Performance vs. Rubin Ultra NVL576 and Groq LPU35 – Disaggregated Inference

Prodigy, Rubin Ultra Inference Performance Normalized to Groq LP35



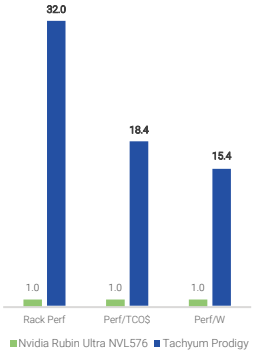
Prodigy, Rubin Ultra Maximum Rack Memory Capacity Normalized to LP35



Prodigy provides up to **488x higher AI Rack Performance** than Groq LP35

HPC Rack Performance vs. Nvidia Rubin Ultra

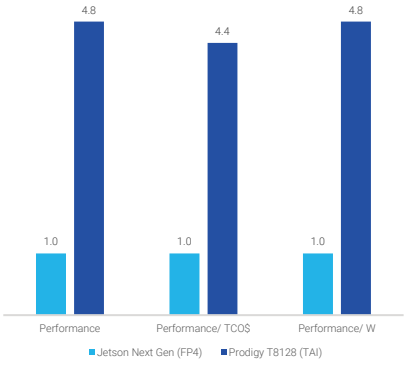
Prodigy vs. Rubin Ultra FP64



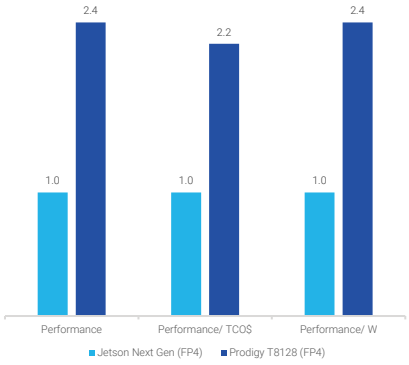
Prodigy provides up to **32x higher HPC Rack Performance** than Rubin Ultra NVL576

Prodigy vs. Nvidia Jetson Next Generation*

Prodigy (TAI) vs. Jetson Next Gen (FP4)



Prodigy vs. Jetson Next Gen (FP4)

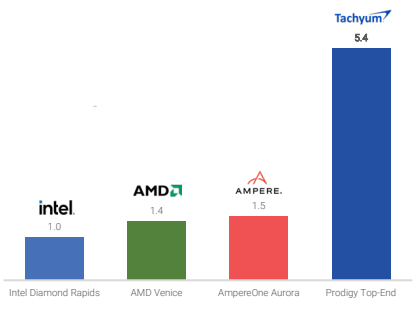


Prodigy provides up to **4.8x higher Performance** and **Performance/W** than Next Generation Jetson

*Jetson next generation projected based on available data

Cloud Performance vs. x86 and Arm

SPECrate 2017 Integer



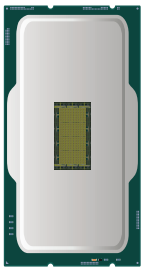
Prodigy Delivers up to **5x Higher Cloud Performance** than x86

Notes: Rubin Ultra and LP35 integrate fixed-sized memory.
Prodigy uses external memory that can be scaled based on customer requirements

Device Architecture

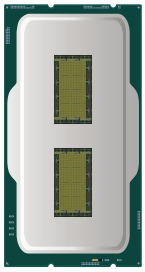
Prodigy's groundbreaking universal processor architecture includes 1024 high-performance CPU cores, 24 DDR5-17600 memory controllers, and 128 lanes of PCIe 7.0 delivering processing power and high memory and I/O bandwidth that is designed to be balanced, optimizing system performance and avoiding bottlenecks. Prodigy's advanced chiplet architecture enables a wide array of scalability from 1024 cores in our premium package down to 32 cores in our entry/mid-range package. This allows Prodigy to excel in a wide range of applications from the highest performance exascale supercomputers to cost and power-efficient physical AI applications such as drones, autonomous vehicles, and humanoid robots. Prodigy will be manufactured in 2nm process technology.

Entry-Level/ Physical AI

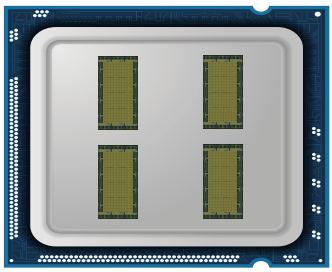


- 120mm x 63mm
- Air-Cooled

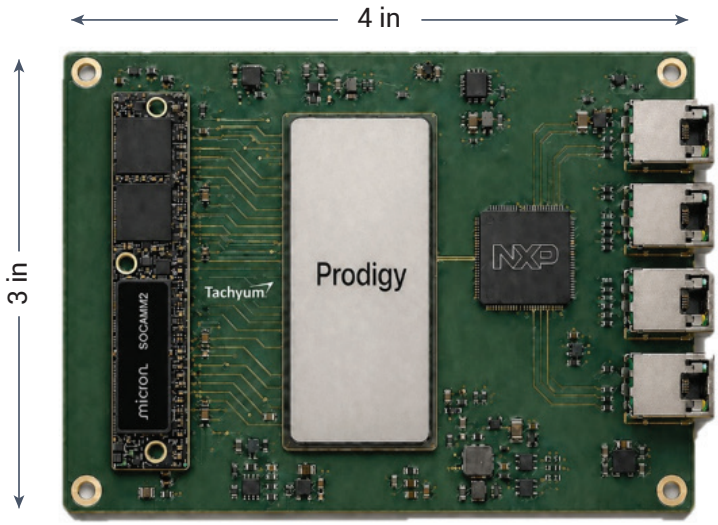
Mid-Range



High-End



- 120mm x 150mm
- Liquid-Cooled



Physical AI

The Prodigy Physical AI Module will deliver high performance and low power for today's demanding AI applications in a compact form factor, with a mock-up of the module shown.

The module supports 32, 64, 96, and 128-core Prodigy Universal Processors for a wide range of performance and power. The module provides up to 38 PF of AI performance and power consumption ranges from 50 to 200 W. The compact module measures only 3 inches x 4 inches. Additional features include a field-upgradable SOCAMM2 module supporting up to 256 GB of LPDDR5x memory running up to 9600 MT/sec, and 4 x 100 GbE networking ports.

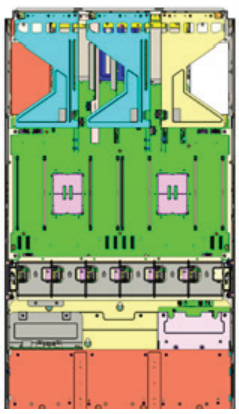
Platform Strategy

Prodigy's platform strategy includes two types of evaluation platforms. A standard, air-cooled 2-socket platform will be used for cloud and AI workloads, and for lead customers who require the highest performance, there is a liquid-cooled 4-socket platform that will be targeted for maximum AI and HPC performance. PCIe slots on both platforms support standard and OCP form factors.

The platforms support simple, out-of-the-box evaluation with an SDK that includes Linux, gcc compiler, software libraries, and a large ecosystem of recompiled native applications, streamlining software development.

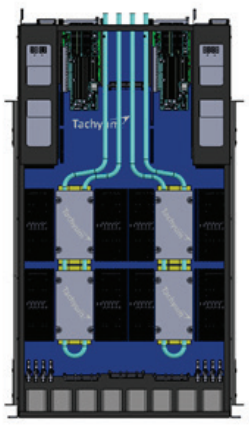
The evaluation platforms provide early customer access, and enable ODM/OEM partners to leverage the platforms to develop production designs and platforms.

Standard Platform for Cloud/AI



Air-Cooled 2-Socket Evaluation Platform

Lead Customer Platform for Maximum AI/HPC Performance



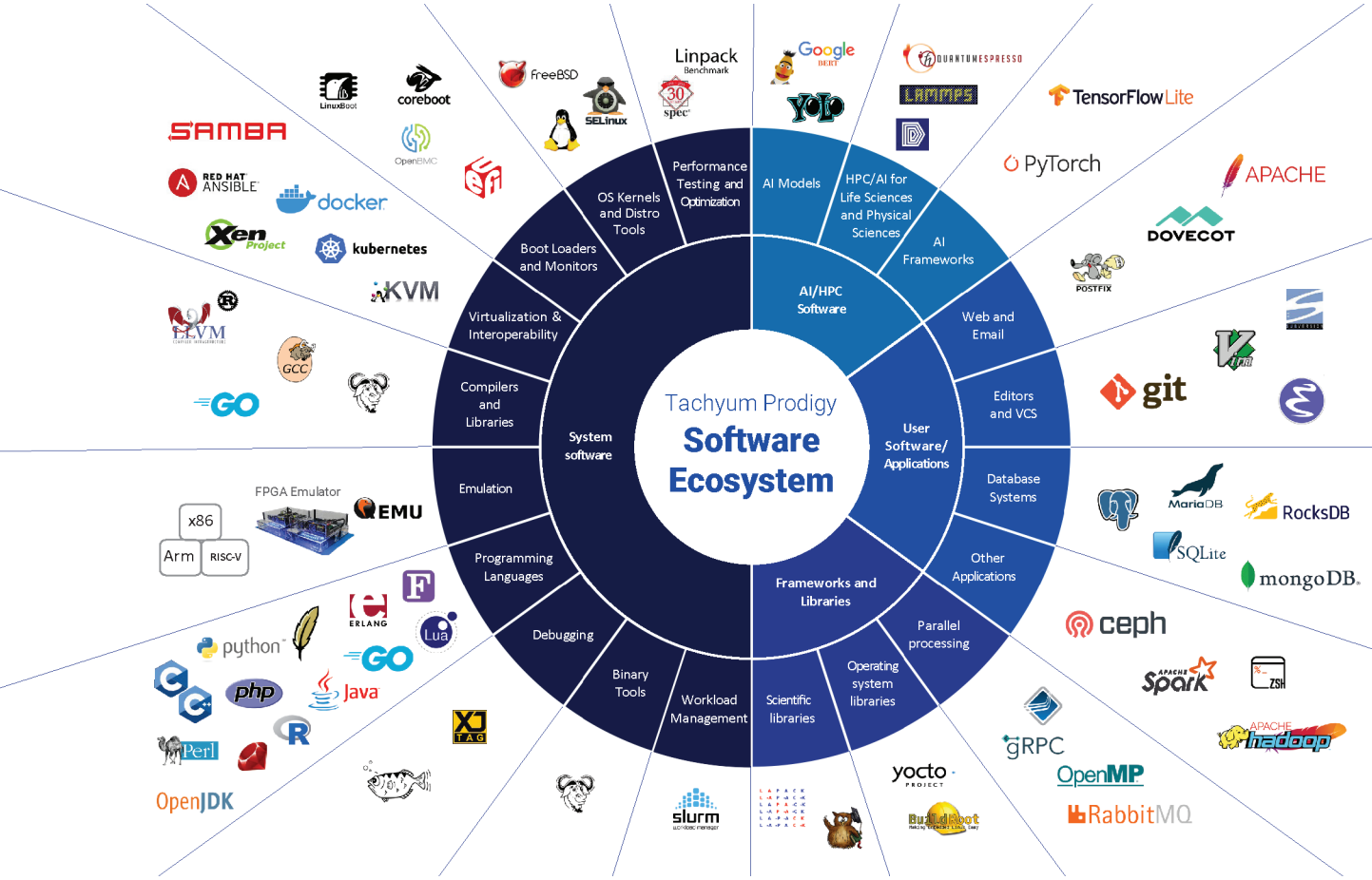
Liquid-Cooled 4-Socket Evaluation Platform

Prodigy SKU Summary

Product SKU	Number of Cores	Max Frequency (GHz)	Scalability	DDR5 Controllers	DDR5 Speed	PCIe 7.0 Lanes	TDP	Target Markets / Applications
T241024	1024	6.0	8S	24	17600	128	1600	Top-End HPC, Big AI
T24768	768	5.0	8S	24	17600	128	1000	Entry HPC, Big AI, Crypto, Digital Currency
T16512	512	6.0	16S	16	12800	128	800	Analytics, Big Data, Crypto, Digital Currency
T16448	448	5.5	16S	16	12800	128	645	Analytics, Big Data, In-Memory Databases
T16384	384	5.5	2S	16	12800	128	550	Cloud, Analytics, Big Data
T16320	320	5.0	2S	16	10700	128	420	Cloud, Edge/Telco
T16256	256	4.5	2S	16	10700	128	300	Cloud, Databases, Edge/Telco
T8256	256	4.5	1S	8	10700	96	300	Cloud, Databases, Edge/Telco
T8128	128	4.5	1S	8	9600	96	150	Physical AI, Cloud, Databases, Storage
T896F	96	5.5	1S	8	9600	96	140	Cloud, Databases, Storage
T896	96	4.0	1S	8	9600	96	100	Physical AI, Cloud, Databases, Storage
T464	64	4.0	1S	4	6400	48	70	Physical AI, Entry Cloud, Storage
T432	32	3.5	1S	4	6400	24	30	Physical AI, Low Power, Hosting

Prodigy Software Ecosystem

Prodigy has a rich ecosystem of development tools, operating systems, application software, and software libraries to enable fast, easy development and quick time to market.



Complete Software Ecosystem at www.tachyum.com/sw

Tachyum

www.tachyum.com

Tachyum Inc., 8275 South Eastern Ave, Ste 233, Las Vegas, NV 89123, U.S.A.
Tachyum s.r.o., Mostová 4, 811 02 Bratislava, Slovakia

Scan the QR Code to discover our products

Brochure 2nm_260609